

REMARKS

The claims have been amended by rewriting claims 7 and 8, canceling no claims, and adding no new claims. Claims 1-14 remain in the application.

Reconsideration of this application is respectfully requested.

Informalities in the Abstract

Applicant thanks the Examiner for the remarks concerning the Abstract. In this instance, the applicant believes that the figure selected for the cover page is likely to be FIG. 2, which matches the numbers in the Abstract. Since the Abstract for this Application is not lengthy, the applicant has chosen not to make the changes. If the examiner indicates an intention to select FIG. 1, the applicant will remove the references.

Claim Rejections - 35 U.S.C. § 112, second paragraph:

Claim 8 was rejected under 35 U.S.C. § 112, second paragraph as being indefinite for failing to particularly point and distinctly the subject matter which the applicant regards as the invention.

The applicant has revised claim 8 by making it dependent upon claim 7, which the applicant believes overcomes the Examiner's rejection.

Claim Rejections - 35 U.S.C. § 102(b):

Claims 1, 4, 5, 9, and 14 (claim 14 is not in Examiner's list at section 6 of the rejection, but is included in Examiner's text) were rejected under 35 U.S.C. § 102(b) as being clearly anticipated by Basu [US Patent 6,594,629.

Applicant respectfully traverses the Examiner's rejection of claims 1, 4, 5, 9, and 14 as being clearly anticipated by Basu. Applicant believes that the Examiner has mischaracterized Basu.

The Examiner states that the following aspect of claim 14 is described by Basu, and in particular at the location cited:

"generate a set of visemes [at column 30, lines 13-28, as subsequently use a phoneme to viseme mapping]." The full wording of Applicant's last element is "analyzing each of the time domain classification vectors to synchronously generate a set of visemes corresponding to each of the successive frames of digitized speech information at the fixed rate", which, in combination with the rest of claim 1, clearly describes generating visemes from audio speech information, not video speech information. In contrast, Basu only describes generating visemes from video information. This is made clear by studying Basu in its entirety. However, conclusive evidence to this affect is given from several sections of Basu. First, here is a section of the document from that cited by the Examiner [which Applicant believes is at col. 20, not col. 30] - Col. 20, lines 12-14: "For video, we experiment with both the phonetic classification and a 'viseme' based approach as described above. One approach to labeling the video feature vectors is to label the speech data from a Viterbi alignment and to subsequently use a phoneme to viseme mapping." The video approach is described above in Basu at col. 13, lines 35-57. Within this section, lines 40-44 specifically describe generating speech data (phonemes) from the video data: "The extracted visual speech feature vectors are then normalized in block 24 with respect to the frontal pose estimates generated by the detection module 20. The normalized visual speech feature vectors are then provided to a probability module 26. Similar to the probability module 16 in the audio information path which labels the acoustic feature vectors with one or more phonemes, the probability module 26 labels the extracted visual speech vectors with one or more previously stored phonemes." ... At col. 13, lines 54-57, "Alternatively, the visual speech feature vectors may be labeled with visemes which, as previously mentioned, are visual phonemes or canonical mouth shapes that accompany speech utterances." Applicant finds no reference in Basu that describes generating visemes from audio speech vectors, which is required by the elements of applicant's claim 1. For reference, the word "viseme" appears in Basu's summary and detailed descriptions at: Col. 2, line 61; Col. 13, line 55; Col. 16, line 37; Col. 16, line 39; Col. 17, line 55; Col. 20, line 13-45.

For these reasons, applicant believes that claims 1a n 14 are patentable over Basu and any combination of Basu and the art cited in this application.

Applicant believes that claims 4, 5, and 9 are patentable because they are dependent upon claim 1, which applicant believes is patentable.

Claim Rejections - 35 U.S.C. § 102(e):

Claims 1, 2, 9, 10, and 12-14 were rejected under 35 U.S.C. § 102(e) as being clearly anticipated by Sutton [US Patent 6,539,354].

Applicant respectfully traverses the Examiner's rejection of claims 1, 2, 9, 10, and 12-14 as being clearly anticipated by Sutton. Applicant believes that the Examiner has mischaracterized Sutton. Sutton, at col. 19, lines 1-13 characterizes the phoneme generation process as follows:

Referring to FIG. 8, a speech input stream 2 B, or speech wave, is received into the system in 10 ms frames at a sampling rate of typically between 8 kHz to 45 kHz (depending on the system capability and the desired speech quality). A feature representation is computed for each frame and assembled into a content (feature) window 6 . The feature window 6 contains 160 ms of speech information or, in other words, data from sixteen 10 ms frames. The feature window 6 is transmitted to a phonetic (phoneme) estimator 10 B. The phoneme estimator 10 B includes a phoneme neural network 16 B which receives the feature window 6 as an input and produces context- dependent phoneme (phone) estimates 12 as an output. The phoneme estimates 12 are then sent to a viseme estimator 30 B.

From this explanation, it can be seen that Sutton generates a plurality of vectors per frame (vectors are generated at 10 msec intervals, frames are 160 msec long; thus there are 16 vectors per frame. In contrast, Applicant's claim 1 includes the description; "each of the time domain frame classification vectors is derived from one of the successive frames of digitized analog speech information".

For these reasons, applicant believes that claims 1 and 14 are patentable over Sutton and any combination of Sutton and the art cited in this application.

Applicant believes that claims 2, 9, and 10 are patentable because they are dependent upon claim 1, which applicant believes is patentable.

Applicant believes that claims 12 and 13 are patentable for the same reasons as claim 1.

Claim Rejections - 35 U.S.C. § 103:

Claims 6-8 were rejected under 35 U.S.C. § 103 as being unpatentable over Basu (US Patent 6,594,629) in view of David J. Thomson, "An Overview of Multiple-Window and Quadratic-Inverse Spectrum Estimation Methods," IEEE 1994, pp. VI 185-VI 194.

Applicant believes that claims 6-8 are patentable because combining Thomson with Basu fails because Basu fails for the reasons described above with reference to the rejection of Claims 1 and 14 over Basu.

Notwithstanding these reasons, applicant believes that claims 6-8 are patentable on their own merits, and respectfully traverses Examiner's rejection thereof, for the reason that Thomson does not teach the motivation of using N MTDPSB functions for which they are used in the claimed invention, which is to achieve the synchronization aspect of claim 1. In fact, the advantage cited by the Examiner of using MTDPSB functions, which is the advantage of achieving the best possible leakage properties for a dynamic range, is purposefully sacrificed in order to perform the computation of the MTDPSB with the low latency needed to achieve synchronization. This is accomplished by using a low value of N.

Claim Rejections - 35 U.S.C. § 103:

Claims 3 and 11 were rejected under 35 U.S.C. § 103 as being unpatentable over Sutton (US Patent 6,539,354) in view of Peterson et al. [US Patent 5,067,095].

Applicant believes that claims 3 and 11 are patentable because combining Sutton with Peterson fails because Sutton fails for the reasons described above with reference to the rejection of Claims 1 and 14 over Sutton.

Notwithstanding these reasons, applicant believes that claims 3 and 11 are patentable on their own merits, and respectfully traverses Examiner's rejection thereof. Applicant believes that the combination of Peterson and Sutton do not show the use of a neural network as

DOCKET NO.: CMLD1339T

claimed by applicant in claim 11, and could not provide a latency of less than 10 msec between an audio input and viseme selection, as suggested by the Examiner. The reasons are 1) there is no obvious way to combine Sutton and Peterson to provide the elements of applicants' claimed invention because there is no description in Peterson of how the neural network would operate with the inputs to the neural network of applicants' claims, and 2) the 0.2 msec latency that is described at the cited portion of Peterson is a latency between an audio signal input and a primitive sound determination. Another level of neural networks (module 14) is required in Petersons to determine a phoneme, and then a viseme has to be determined from one or more phonemes, which may require yet another level. Peterson does not describe the number of delays needed for module 14, and does not suggest how visemes would be selected.

Other changes

Claim 7 has been changed to delete obvious redundant language.

Accordingly, this application is believed to be in proper form for allowance and an early notice of allowance is respectfully requested.

Please charge any fees associated herewith, including extension of time fees, to 502117.

SEND CORRESPONDENCE TO:

Motorola, Inc.
Law Department

Customer Number: 22917

Respectfully submitted,

By:


James A. Lamb

Attorney of Record

Reg. No.: 38,529

Telephone: (847) 576-5054

Fax No.: (847) 576-3750

Email: jim.lamb@motorola.com